

УДК 343.98+543.421/.424

**А. Н. Хох**

заведующий лабораторией исследования материалов, веществ и изделий,  
НПЦ Государственного комитета судебных экспертиз Республики Беларусь  
г. Минск, Беларусь

E-mail: [1ann1hoh@gmail.com](mailto:1ann1hoh@gmail.com)

**В. Б. Звягинцев**

*кандидат биологических наук, доцент*  
заведующий кафедрой лесозащиты и древесиноведения,  
Белорусский государственный технологический университет  
г. Минск, Беларусь

E-mail: [mycolog@tut.by](mailto:mycolog@tut.by)

## ОСОБЕННОСТИ ИДЕНТИФИКАЦИИ «СХОЖИХ» ДРЕВЕСНЫХ ПОРОД С ПОМОЩЬЮ ХЕМОМЕТРИЧЕСКИХ АЛГОРИТМОВ

В данной работе показаны возможности использования спектроскопии в ближней инфракрасной области и хемометрической идентификации. Проиллюстрировано применение метода главных компонент, дискриминантного анализа проекций на латентные структуры. Приведены практические примеры использования рассмотренных хемометрических алгоритмов при проведении сравнительного исследования и идентификации «схожих» древесных пород. В целом, оценка предсказательных качеств полученных моделей показала их стабильность. Таким образом, при анализе неизвестного образца древесины попадание его спектра в отдельный кластер свидетельствует о его принадлежности древесной породе, аналогичной другим образцам этого кластера.

**Ключевые слова:** древесина; БИК-спектроскопия; видовая принадлежность; метод главных компонент; дискриминантный анализ; хемометрические алгоритмы.

**A. N. Khokh**

SPC of the State Forensic Examination Committee of the Republic of Belarus  
Minsk, Belarus

**V. B. Zviagintsev**

*Ph.D. (Biology Sciences), Associate Professor*  
Belarusian State Technological University  
Minsk, Belarus

## IDENTIFICATION PECULIARITIES OF "SIMILAR" TREE SPECIES USING CHEMOMETRIC ALGORITHMS

The paper shows the possibilities of using spectroscopy in the near-infrared region along with chemometric identification. The authors have illustrated practice of employing the method of principal components discriminant analysis of projections on latent structures. They have demonstrated the practical examples of the considered chemometric algorithms usage while carrying out the age comparative study and identification of "similar" tree species. In total, the assessment of the obtained models predictive qualities

showed their stability. Thus, when analyzing an unknown wood sample, its spectrum falling into a separate cluster, indicates that it belongs to a tree species similar to other samples of this cluster.

**Keywords:** wood; NIR spectroscopy; species identification; principal component analysis; discriminant analysis; chemometric algorithms.

Древесина относится к наиболее распространенным материалам, изделия из которых окружают человека на протяжении всей жизни. Во многом именно этим объясняется ее частая встречаемость среди вещественных доказательств при расследовании многих преступлений: убийств, поджогов, краж, дорожно-транспортных происшествий, незаконных рубок. При этом следует отметить, что практически в 100% случаев при проведении экспертного исследования требуется установить видовую принадлежность (т. е. породу) исследуемых образцов. В большинстве случаев решение данной задачи проводится на основании анализа макро- и микроскопических признаков и не представляет серьезных трудностей в практическом плане.

Однако в случае предоставления на исследование мелких фрагментов и частиц древесины (щепы, стружка, опилки и др.) могут возникать ошибки, обусловленные отсутствием достаточного комплекса диагностических признаков, позволяющих с уверенностью отнести исследуемый образец к конкретной породе. В основном это характерно для древесных пород с чрезвычайно схожим строением: сосна и лиственница; сосна и пихта; каштан и дуб; липа и осина; и др. В таких случаях возникает необходимость в использовании других методов исследования, которые позволили бы отличить «схожие» древесные породы и получить максимум информации диагностического и идентификационного характера. К таким методам на сегодняшний день можно отнести неразрушающие спектроскопические методы анализа, среди которых спектроскопия в ближней инфракрасной области (далее – БИК-спектроскопия) [1–4]. Следует отметить, что по результатам проведенных нами ранее исследований уже было показано, что данный метод может быть эффективно использован для решения экспертных задач, связанных с установлением места происхождения сосновых лесоматериалов [5], а также идентификацией целого по частям [6].

Цель работы – разработать подходы для дифференциации «схожих» пород древесины на основе спектральной информации.

Объекты исследования – точно идентифицированные до вида образцы ели обыкновенной (*Picea abies*) и пихты сибирской (*Abies sibirica*) – пара № 1; клена остролистного (*Acer platanoides*) и клена ложноплатанового (*Acer pseudoplatanus*) – пара № 2; тополя белого (*Populus alba*) и осины обыкновенной (*Populus tremula*) – пара № 3 из ксилотеки кафедры лесозащиты и древесиноведения Белорусского государственного технологического университета [7] и натурной коллекции НПЦ Государственного комитета судебных экспертиз Республики Беларусь. Для проведения экспериментальных работ была подготовлена репрезентативная выборка образцов, которая учитывала различные участки дерева. Количество исследованных образцов для каждой породы составило от 20 до 30. Всего проанализировано 129 образцов.

БИК-спектры регистрировали на спектрометре MicroNIR OnSite (VIAVI, США) в диапазоне 950–1650 нм без предварительной специальной пробоподготовки. Для каждого образца на тангенциальном срезе древесины проводили по 20 последовательных измерений, которые в дальнейшем усреднялись до одного спектра. Статистическую обработку полученных спектров проводили с использованием пакета программ The Unscrambler X v.10.4.1 (CAMO, США). Для оценки полученных результатов были применены методы главных компонент (далее – PCA) и дискриминантного анализа проекций на латентные структуры (далее – PLS-DA).

В результате проведенных исследований было установлено, что по абсолютной величине БИК-спектры 3-х пар «схожих» древесных пород имеют схожий вид и несущественные различия (рисунок 1).

В связи с этим нами проводилось вычисление производной второго порядка по методу Савицкого-Голея (с использованием 7 сглажи-

вающих точек), что позволило сделать признаки разделения (так называемые индивидуализирующие признаки) более явными. Однако, в любом случае, существенных изменений в интенсивности БИК-спектров между образцами древесины из пар № 1–3 не обнаружено,

что обусловило необходимость применения хемометрических алгоритмов анализа.

Первоначально спектрометрические данные для каждой из 3-х исследованных пар древесных пород были изучены методом РСА (рисунок 2).

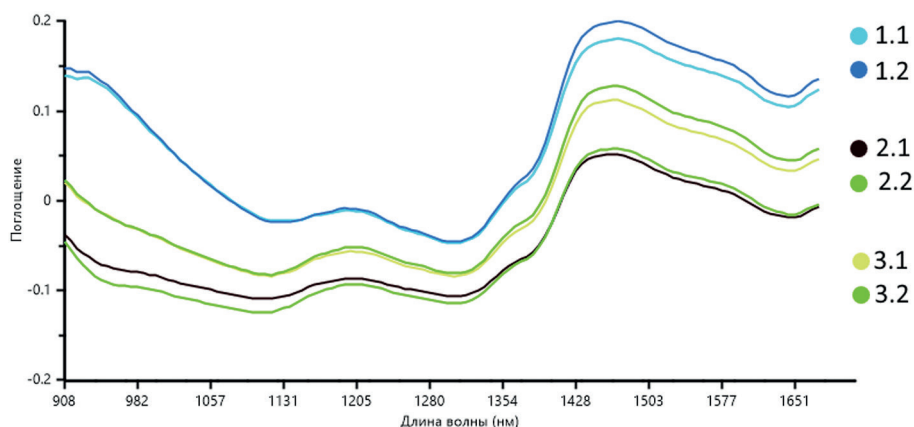
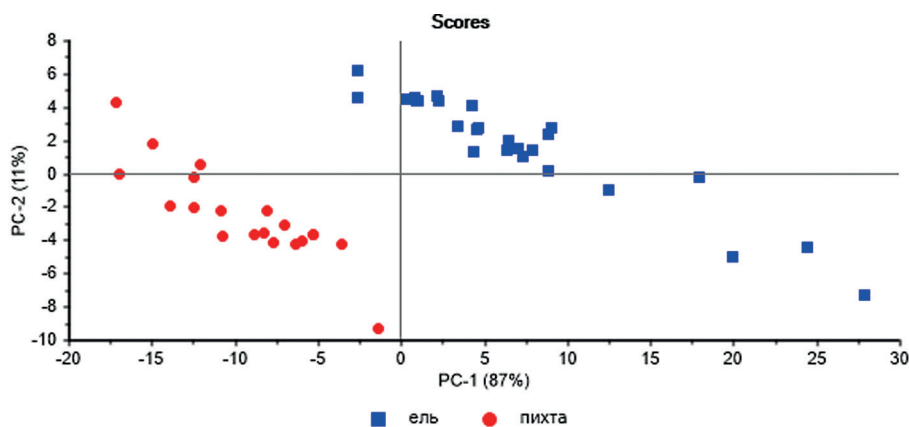
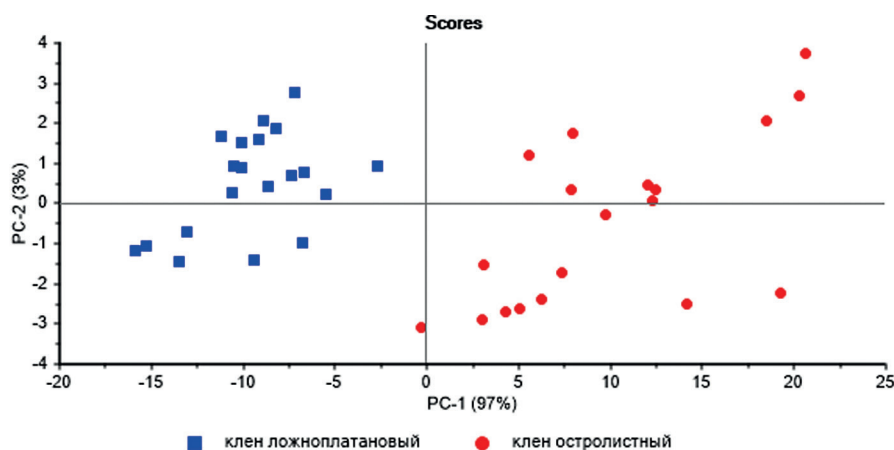


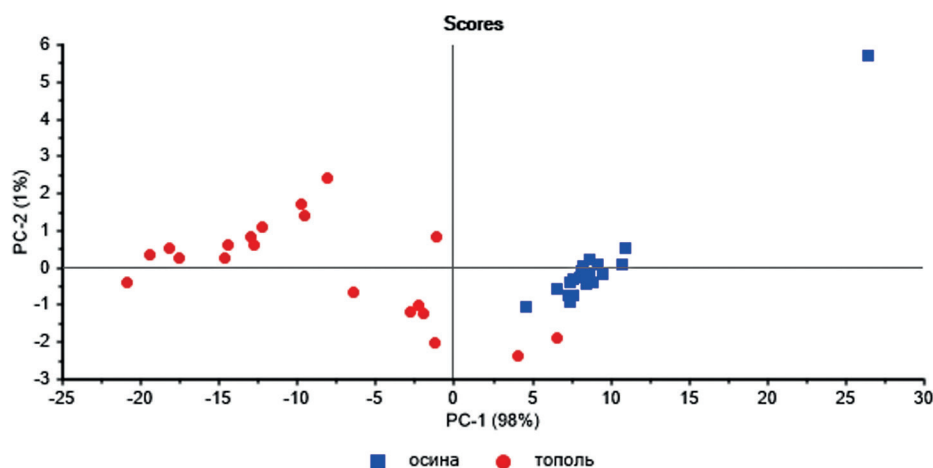
Рисунок 1 – Усредненные абсолютные БИК-спектры для исследованных пород: 1.1 – ель обыкновенная; 1.2 – пихта сибирская; 2.1 – клен остролистный; 2.2 – клен ложноплатановый; 3.1 – тополь белый; 3.2 – осина обыкновенная



а



б



6

Рисунок 2 – Графики МГК-счетов для исследованных образцов

На основании проведенного анализа можно видеть, что точки, являющиеся отражением исследованных образцов, в каждом из 3-х рассматриваемых случаев объединились в 2 класса, что соответствует 2-м исследуемым породам.

Как видно на рисунке 2 а-в, точки, соответствующие одной породе, практически не перекрываются и находятся в разных областях модели относительно 1-й и 2-й главных компонент (далее – PC). Это свидетельствует о достоверных различиях между ними. Распределение образцов в координатном пространстве PC показывает, что образцы ели обыкновенной, осины обыкновенной и клена ложноплатанового локализованы в зоне отрицательных значений (за исключением 2-х образцов осины), в то время как образцы из пихты сибирской, клена остролистного и тополя белого находятся в зоне положительных значений (за исключением 2-х образцов ели и клена 1-го образца). Следует отметить, что вследствие того, что древесина является естественным объектом, она представляет собой сложный объект для исследования методом БИК-спектроскопии, а потому все исследованные породы характеризуются довольно большим разбросом. Оценка полученных результатов показала, что первая главная компонента объясняет от 87 до 99% дисперсии (в соответствии с поставленной задачей распределение образцов изучено в координатах PC1-PC2), т. е. именно она описывает основные отличия между северными

и южными образцами, при этом данные отличия являются самыми значительными.

Следующий этап работы заключался в непосредственном многомерном моделировании БИК-спектров. С учетом специфики задачи нами было решено использовать метод дискриминантного анализа проекций на латентные структуры, что позволило в дальнейшем предсказывать принадлежность новых образцов. Его суть состоит в том, что дискриминационные правила для классов задаются линейными регрессионными уравнениями. Данный метод относится к методам обучения с учителем. Его основное преимущество заключается в возможности проведения одновременного многоклассового анализа.

Для анализа данных методом PLS-DA использовались те же данные, что и для PCA, однако при этом наблюдения предварительно были разделены на классы, соответствующие исследуемым образцам. Так, ели обыкновенной, клену остролистному и тополю белому был присвоен класс 1; пихте сибирской, клену ложноплатановому и осине обыкновенной – класс -1. В первую очередь были построены классификационные модели для каждого класса в отдельности. Для проверки их работоспособности был выбран метод кросс-валидации. Все образцы класса были 10 раз разделены рандомизированным образом на обучающий (примерно 2/3 образцов) и проверочный (примерно 1/3 образцов) наборы данных. Результаты представлены в таблице 1.

Таблица 1 – Результаты классификации БИК-спектров исследованных образцов (% правильно распознанных образцов)

Набор данных (число образцов)	Число образцов в обучающем и тестовом наборе*	PLS-DA
Ель обыкновенная (n=27)	Обучающий набор (18 образцов)	96
	Тестовый набор (9 образцов)	94
Пихта сибирская (n=20)	Обучающий набор (14 образцов)	94
	Тестовый набор (6 образцов)	92
Клен остролистный (n=20)	Обучающий набор (14 образцов)	94
	Тестовый набор (6 образцов)	92
Клен ложноплатановый (n=20)	Обучающий набор (14 образцов)	92
	Тестовый набор (6 образцов)	91
Осина обыкновенная (n=20)	Обучающий набор (14 образцов)	86
	Тестовый набор (6 образцов)	84
Тополь белый (n=20)	Обучающий набор (14 образцов)	84
	Тестовый набор (6 образцов)	81

\* Разбиение на обучающий/тестовый набор проведено 10 раз случайным образом

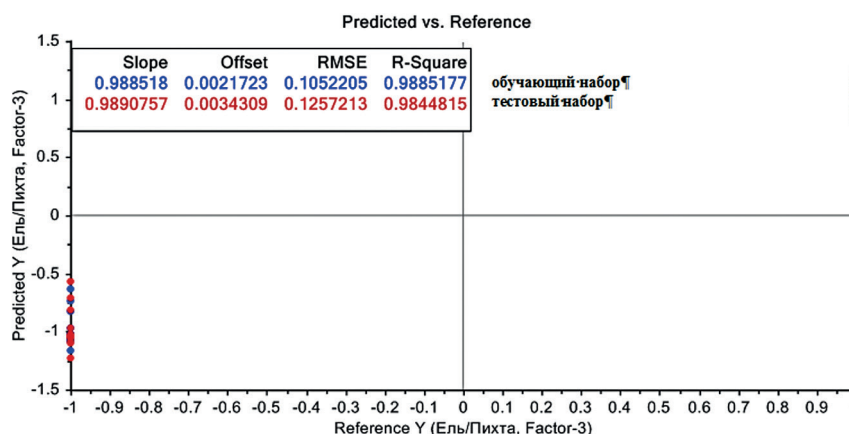
Обучающие наборы образцов использовались для построения PLS-DA модели, с помощью которой в последующем будет проводиться анализ неизвестных экспертных образцов. В результате отмечено, что так же, как и в PCA модели, классы практически не перекрывались между собой и расположились в разных областях двумерной плоскости, что подтверждает различия между ними. При этом отметим, что оптимально число латентных переменных (PLS-DA факторов) соответствовало 3-м для полных и обучающих наборов данных для обоих классов в паре № 1–2 и 6-ти для обоих классов в случае пары № 3.

Суммарный процент объясненной информации практически не отличался от PCA моделей и составил от 97 до 98%. Объясненная дисперсия для PLS-DA модели показывает,

что достаточным является учет первых двух факторов. Средняя достоверность PLS-DA моделей, оцененная по методу перекрестной проверки, составила 90%.

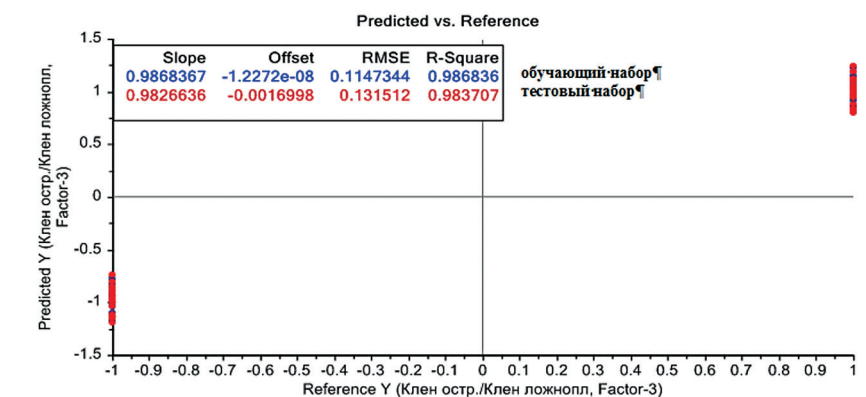
Практические возможности использования созданной классификационной модели были проверены на реальных объектах, поступающих для проведения судебно-экспертных исследований.

Анализ результатов методом PLS-DA не показал существенного различия в значениях указанных характеристик для калибровки и предсказания для ели и пихты, 2-х видов клена, что предполагает низкую погрешность в дальнейших определениях. В то время как для тополя и осины наблюдались отличия, которые свидетельствует о недостаточной стабильности модели, однако не исключают возможность ее использования (рисунок 3).

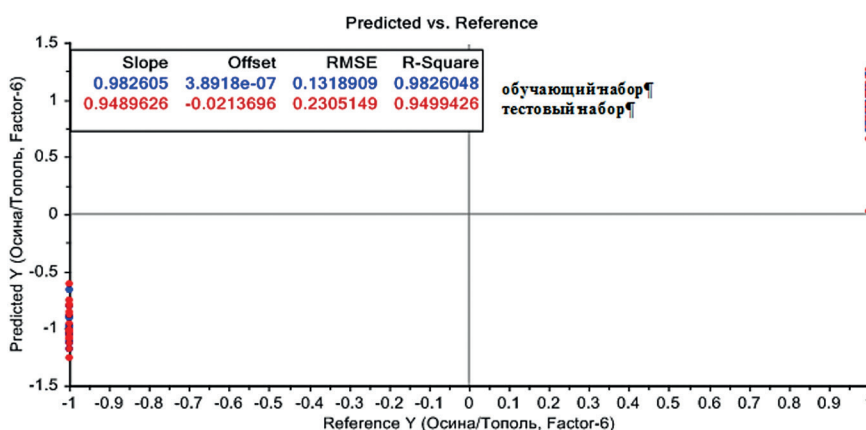


a





б



в

Рисунок 3 - Определение видовой принадлежности (проверка – методом полной кросс-валидации)

В таблицах 2–4 показаны коэффициенты значимости идентифицированных экспертных образцов в формировании осей Factor-1 и Factor-2 и расположении точек описанных выше моделей. Указанные коэффициенты име-

ют как положительные, так и отрицательные значения в зависимости от их вклада в положительную или отрицательную области модели, их значимость при этом не зависит от знака и определяется только отклонением от нуля.

Таблица 2 – Коэффициенты значимости идентифицированных неизвестных образцов в формировании осей Factor-1 и Factor-2

Неизвестные образцы (Ель/Пихта, Factor-3)	Прогнозируемое значение	Отклонение
1	0.9675095	0.04956548
2	0.9967663	0.05716535
3	1.003229	0.06243287
4	1.02304	0.06212076
5	1.019375	0.07706694
6	-1.09266	0.1819288
7	-1.017764	0.1449992
8	-1.167552	0.2775553
9	-1.02475	0.3514048
10	-0.9802454	0.2045099

Таблица 3 – Коэффициенты значимости идентифицированных неизвестных образцов в формировании осей Factor-1 и Factor-2

Неизвестные образцы (Клен остр./Клен ложнопл., Factor-3)	Прогнозируемое значение	Отклонение
1	0.9363644	0.1254152
2	0.9150217	0.09262475
3	1.03194	0.07723044
4	-0.9076848	0.1890156
5	-0.8201427	0.2375492
6	-0.8831725	0.1900632

Таблица 4 – Коэффициенты значимости идентифицированных неизвестных образцов в формировании осей Factor-1 и Factor-2

Неизвестные образцы (Осина/Тополь, Factor-6)	Прогнозируемое значение	Отклонение
1	1.131052	0.1750191
2	0.9181299	0.296646
3	1.040781	0.1769006
4	-0.9950018	0.3337322
5	-1.141456	0.4554893
6	-1.04321	0.2163651

Таким образом, с учетом рассчитанных коэффициентов значимости, исследованные неизвестные образцы были отнесены к конкретной породе. Отметим, что их видовая принадлежность была подтверждена с помощью микроскопического исследования. В результате все неизвестные образцы с использова-

нием 3-х построенных нами моделей были классифицированы правильно. Так, образцы № 1–5 в модели ель/пихта отнесены к породе ель, а образцы № 6–10 – пихта (рисунок 4). Результаты по 2-м другим моделям представлены на рисунках 5–6.

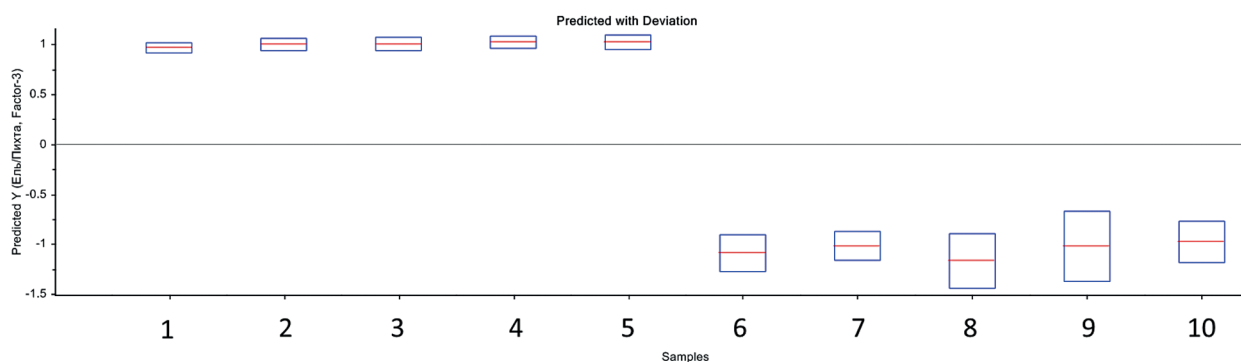


Рисунок 4 – Разделение неизвестных образцов: 1 – ель; -1 – пихта

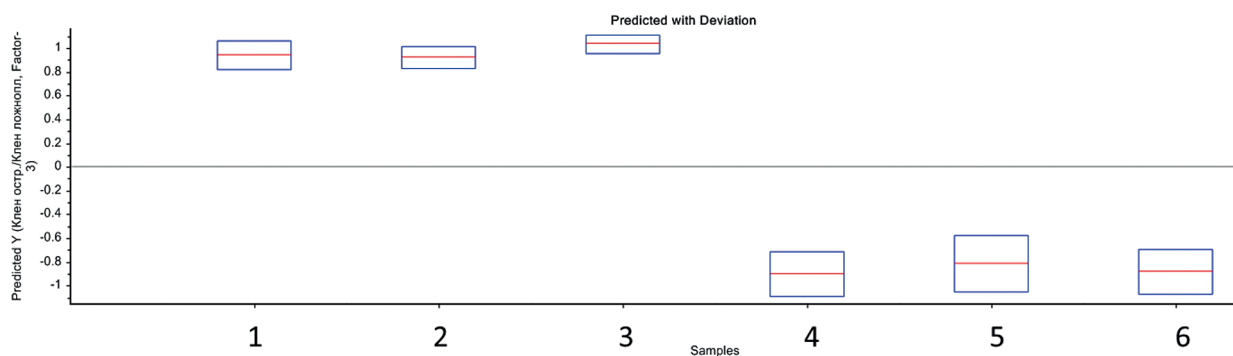


Рисунок 5 – Разделение неизвестных образцов: 1 – клен остролистный; -1 – клен ложноплатановый

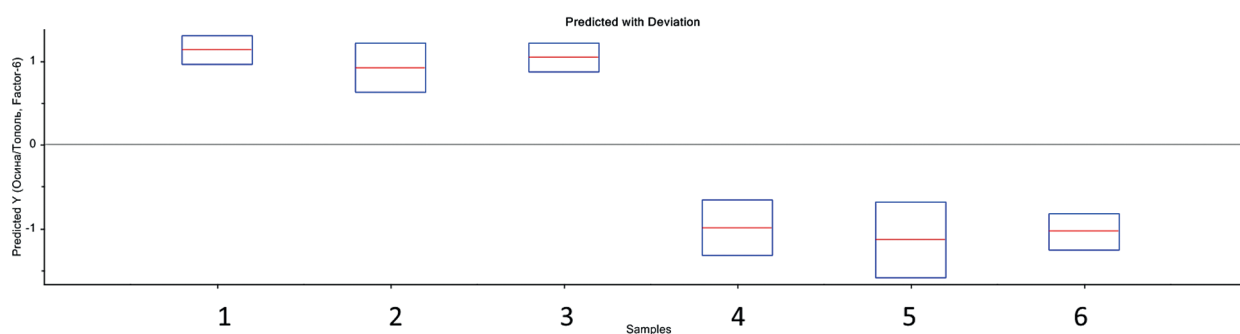


Рисунок 6 – Разделение неизвестных образцов: 1 – осина; -1 – тополь

Таким образом, полученные для проанализированных с использованием хемометрических алгоритмов классификации БИК-спектров результаты показали, что «схожие» древесные породы могут быть успешно классифицированы с использованием дискриминантного анализа проекций на латентные структуры. Необходимым условием при этом является создание постоянно обновляемой базы БИК-спектров, характерных для конкретной породы древесины. Применение PLS-DA позволяет сравнивать неизвестные образцы

с контрольными, существующими в базе данных, и на этой основе классифицировать новые образцы, поступающие для проведения экспертных исследований. В целом можно констатировать, что метод БИК-спектроскопии подходит для установления видовой принадлежности древесины, а потому может быть рекомендован к применению в судебно-экспертной практике. При этом использование хемометрических алгоритмов позволяет существенно упростить и ускорить проведение экспертизы.

### Список использованных источников

1. Classification of red oak (*Quercus rubra*) and white oak (*Quercus alba*) wood using a near infrared spectrometer and soft independent modelling of class analogies / O. E. Adedipe [et al.] // *Journal of Near Infrared Spectroscopy*. – 2008. – V. 16. – № 1. – P. 49–57.
2. Cognitive spectroscopy for wood species identification: near infrared hyperspectral imaging combined with convolutional neural networks / H. Kanayama [et al.] // *Analyst*. – 2019. – V. 144. – № 21. – P. 6438–6446.
3. Lazarescu, C. Wood species identification by near-infrared spectroscopy / C. Lazarescu // *International Wood Products Journal*. – 2017. – V. 8. – № 1. – P. 32–35.
4. Ziyang, W. A. N. G. Identification of common wood species in northeast China using Vis/NIR spectroscopy / W. A. N. G. Ziyang // *Journal of Zhejiang A&F University*. – 2019. – V. 36. – № 1. – P. 162–169.



5. Хох, А. Н. Установление географического происхождения сосны обыкновенной (*Pinus sylvestris* L.) по спектрометрическим параметрам в БИК-диапазоне / А. Н. Хох, В. Б. Звягинцев / Весн. Віцеб. дзярж. ўн-та. – 2020. – № 1 (106). – С. 78–84.
6. Хох, А. Н. Пути совершенствования судебно-ботанической экспертизы в Республике Беларусь / А. Н. Хох, В. Б. Звягинцев // Судебная экспертиза Беларуси. – 2019. – № 2. – С. 67–72.
7. Звягинцев, В. Б. Научная и образовательная роль Ксилотеки БГТУ / В. Б. Звягинцев, О. И. Войнич, В. А. Ярмолович // Высшее техническое образование: науч.-метод. журн. – Минск, 2019. – С. 53–58.

Дата поступления: 30.09.2020